OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

OeNB Financial Literacy Evaluation Series 2025/06
20 pages

Conducting representative surveys and impact assessments: an overview of survey methodology

In this paper, we provide an accessible introduction to survey methodology in the context of financial literacy research. We explore why and how well-designed survey methods are necessary for improving the accuracy and quality of insights about larger populations. We cover key methodological basics, particularly sampling strategies and related methods.

Authors

Maximilian Zieser
OeNB, Financial Literacy and Culture Division
maximilian.zieser@oenb.at

Theresa Lorenz
OeNB, Financial Literacy and Culture Division
theresa.lorenz@oenb.at

Valentin Voith
OeNB, Financial Literacy and Culture Division
valentin.voith@oenb.at

What is survey methodology?

Survey methodology is the combination of methods used to design and conduct surveys that allow accurate conclusions about a population. In financial literacy research, using appropriate methods is essential to ensure realistic and accurate results.

How to tame your bias

Survey results can be biased (i.e. distorted) when parts of the population are more likely to be included than others. Balancing practicality and accuracy, researchers can reduce bias through careful and deliberate sampling, contact strategies and weighting.

Sampling matters

A well-chosen sample is the foundation for meaningful results. Researchers need this foundation to estimate statistics (like average financial knowledge) that are valid beyond the surveyed group, especially when combined with appropriate weighting and recruitment practices.

Abstract

Yielding findings from samples that are applicable to broader populations is a fundamental goal in financial literacy research. Generating valid and representative findings is methodologically challenging, however. It requires the careful consideration of sampling strategies, survey modes, participant recruitment and weighting procedures – methods that are often summarized under the umbrella term of survey methodology. This paper offers an accessible introduction to the core principles of survey methodology, with a particular focus on how these principles can support robust, representative insights gained from financial literacy research. It aims to introduce financial literacy program designers to these fundamental methodological considerations and encourage researchers to conduct more robust and methodologically transparent studies and evaluations.

Introduction

Quantitative research usually aims for results that are applicable to a wider population. In the context of financial literacy, this may concern e.g. nation-wide surveys on the levels of financial literacy and similar properties of a country's adult population or the effects of a financial education intervention on the saving behavior of a country's secondary level students. Obtaining results that can be generalized to the target populations is thus a fundamental aspect of research on financial literacy and financial education.

Imagine a research project designed to assess the level of financial literacy of a country's adult population. To accomplish this, researchers could develop a survey that measures key dimensions such as financial knowledge, attitudes and behaviors. To ensure that their findings are not merely reflective of those who responded to the survey but representative of the broader population, the survey must be grounded in robust principles of *survey methodology*.

Survey methodology is a collection of methodological concepts and approaches concerned with enabling valid and accurate statistical inferences from samples conducted through surveys. Often, however, the foundations of survey methodology, such as sampling, take a back seat in the design, analysis, interpretation and reporting of studies, despite the undeniable importance of survey methodology to the overall quality of most quantitative research in the field.

This paper explores the foundational concepts of survey methodology. We focus on the aspects of sampling and touch on the closely related issues of participant recruitment, survey modes and weighting. The paper presents conceptual foundations and methodological approaches in an accessible way, aiming to highlight their importance for research in the context of financial literacy and education and to provide financial literacy researchers with the fundamental insights necessary to understand and apply survey methodology.

Because real-world research and survey designs are sometimes complex and strongly depend on the specific research context, this paper is limited to the conceptual foundations. Comprehensive introductory textbooks such as Groves et al. (2004) and Lohr (2021) provide additional details and cover a much wider range of concepts and methods. Methodological reports of large-scale survey or assessment projects (e.g. Albacete et al., 2023; European Social Survey, 2025b; OECD, 2024) provide concrete and detailed examples of complex, rigorous methodological approaches in different research contexts.

In the following sections, we explore basic conceptual foundations of survey methodology: In section 1, we outline the logic behind sampling, including the concepts of populations, samples, representativeness, bias, error and variance. In section 2, we describe common probability and nonprobability sampling approaches. In section 3, we highlight the importance of recruitment strategies and survey modes in fieldwork. In section 4, we briefly outline the logic behind different weighting approaches.

1 Sampling for statistical inference

Random sampling is the foundation of *statistical inference*, i.e. of the process of drawing conclusions about the broader population from a sample (see Lorenz et al., 2025, for a basic introduction to quantitative data analysis including statistical inference). This section covers key concepts of statistical inference closely connected to sampling, including populations, parameters, statistics, error, biases, representativeness, (external and internal) validity and generalizability. It explores how and why drawing random samples is essential for statistical *inferences* in any quantitative research context, whether it is used to compute a simple average, to measure complex constructs (e.g. financial literacy), to assess associations between variables or to conduct randomized experiments.

1.1 Population parameters and sample statistics

The population is the entire group of entities – typically individuals – researchers wish to research and learn about. It is specified by the researchers for their respective research project. Its size can range from very large, e.g. every resident in a country, to very small, e.g. all participants in an education initiative.

Quantitative researchers are often tasked with accurately measuring specific characteristics of an entire population, ranging from simple physical attributes (such as height) to abstract concepts (such as financial knowledge) to relationships between groups or variables, such as differences or correlations. In statistical inference, a numerical value that describes a characteristic of an entire population (e.g. the average age of all adults in a country) is called a *population parameter*. Since in most cases it would be very costly, impractical or even impossible to examine every single entity of a population, researchers often must rely on data from a subset of the population – a sample – to estimate the population parameter. For this purpose, they calculate a *sample statistic*.

For example, researchers may want to calculate the mean (average) from of a financial knowledge test conducted with 1,000 people living in a country (the sample statistic) that reflects the *true* average financial knowledge of all adults living in that country (the population parameter). Moreover, they may also be interested in *differences* in financial knowledge between population groups defined by sociodemographic characteristics such as age or gender. They thus seek difference estimates that reflect the *true* differences in the population. In general terms, any estimated statistic (e.g. a mean, difference or correlation) should accurately reflect the true population parameter.

Box 1

Key terms of statistical inference

Target population: the entire group of entities (typically individuals) that researchers aim to study and learn about. The target population is defined by the researchers and can range in size from small groups (e.g. financial literacy program participants) to large groups (e.g. all residents of a country).

Population parameter: a true characteristic or measure of the entire target population, such as the mean financial knowledge score of all adults in a country. The population parameter is what researchers seek to estimate using data.

Sample: a subset of the target population. A sample is used when studying the entire population is impractical.

Sample statistic: a numerical estimate derived from a sample that is used to approximate a population parameter. Examples of sample statistics are e.g. means, differences and correlations.

1.2 Sampling principles

To perfectly calculate the population parameter of the mean financial knowledge, researchers would need to obtain answers from every single person of the population. As surveying an entire population is often not feasible, it is necessary to select a sample.

The goal of sampling is often (somewhat vaguely) termed "representativeness." However, the concept of representativeness lacks a single, universal definition. For example, it is often linked to how well a sample reflects the population with regard to the distribution of certain population characteristics, e.g. age and gender (see section 2.2.4).

In a general statistical sense, though, a sample can be considered representative of a population if *any* relevant statistics calculated from it closely reflect the population parameters, i.e. the statistics that would be obtained if the *entire* population participated in the respective survey. In other words, a sample can be considered representative when insights derived from it can be generalized to the entire population. This means that there is low *error* that stems from the selection of participants, i.e. the differences between population parameters and statistics are minimized both in terms of *random sampling error* and in terms of systematic *sampling bias*, two concepts that are explored in subsection 1.4. We employ this statistical understanding of representativeness for the remainder of this paper.

The conceptually simplest way to achieve a representative sample is to draw a large, perfectly *random* sample from the whole population. While not every entity of the population is actually represented in the sample, by using such a *simple random sample*, each entity has the *same probability* of being selected into the sample (see subsection 2.1.1).

If every entity that is sampled participates in the survey, increasing the number of randomly sampled entities (i.e. the sample size) improves the chances that any population characteristic is well-represented in the data. Consequently, any sample statistic tends to approach the true population parameter as the sample size grows. Random sampling thus also enables *unbiased* statistical inferences about the target population, with confidence increasing with sample size. In a simple random sample, the sampling bias is (by definition) zero because each participant has an equal chance of being sampled (for an explanation of sampling bias, see subsection 1.3).

All statistics such as means, medians, difference estimates, correlations, regression coefficients and their associated significance tests or confidence intervals thus rely on sampling in the sense outlined above. With a random sample, the only limiting factor for the accuracy of an estimate is sample size. In other words, a sufficiently large simple random sample allows for deriving statistics that reflect any population parameters with high confidence.

Box 2

Representativeness in relation to generalizability, external validity and internal validity

Statistical representativeness is related to other terms that are frequently used in statistics, particularly the concepts of generalizability and external and internal validity. While statistical representativeness refers to the accuracy of a measurement resulting from the sampling process, i.e. to its generalizability to a specific target population, the concept of generalizability alone refers more broadly to the applicability of results, e.g. to other populations or contexts. External validity focuses on the findings of a study and whether they are applicable to contexts outside the research context. A study may have low external validity even if it provides a precise estimate of a population parameter, especially when the research setting – e.g. a highly controlled laboratory experiment – is overly artificial and does not reflect real-world conditions. In such cases, internal validity may still be high if the study accurately captures the causal relationship within that specific context. Internal validity is often ensured through randomization. The relation of randomization to sampling is outlined in subsection 1.5.

In practice, obtaining a simple random sample of a population is often anything but simple. First, sampling requires reliable data on the target population to serve as the *sampling frame*, e.g. a complete address register. From the sampling frame, a random sample can be drawn, which results in the *gross sample*. Second, the sampled entities in the gross sample, e.g. households, must be contacted, invited and willing to participate in the survey. Only then do they become part of the *net sample*, i.e. the final set of entities of which data are included in the analysis. The process of first selecting the sampling frame, then the gross sample and ultimately arriving at the desired net sample comes with many practical obstacles and thus several potential sources of bias (see section 1.5).

There are generalizations of the simple random sample that aim to mitigate some of the practical challenges of sampling from large populations. These sampling methods are no longer "simple" but are considered complex *probability sampling* methods, as they assign different probabilities of selection to different entities (see section 2.1). All sampling approaches that assign known and nonzero probabilities of selection to each unit in the population are referred to as probability samples or as (complex) random samples. They allow for the same valid inferences about the population as simple random samples, provided that proper design weights and variance correction techniques are used (see section 4). However, they come with many of the same challenges in achieving an unbiased gross and net sample as simple random sampling.

Nonprobability sampling, in contrast to probability sampling, typically does not rely on complete sampling frames of the population nor on systematic selection or invitation approaches (see subsection 2.2). It thus constitutes a fundamentally different class of sampling approaches that often severely limit the ability to identify sources of bias or correct bias through weighting and thus to make statistical inferences. In practice, the principles and methods presented in this paper are primarily applicable to probability sampling approaches, where a complete sampling frame exists and the gross and net samples are clearly identified.

Box 3

Key terms of sampling

Statistical representativeness: in the context of statistical inference, the ability of sample statistics to reflect population parameters with low error. The theoretically simplest way to achieve a representative sample is drawing a large simple random sample.
Simple random sample: a sample drawn in such a way that each entity in the population has the same probability of being selected. If all sampled entities participate, this method guarantees that the sampling bias is zero and that the sample statistics will approximate the true population parameters better the larger the sample is.
Sampling frame: a register of entities that reflects the target population and can be used as the basis for sampling, e.g. the list from which individuals are randomly drawn
Gross sample: the sample that results from selecting entities from the sampling frame, e.g. the list of sampled individuals that are invited to participate in the survey
Net sample: the final set of entities from which data has actually been collected, e.g. individuals whose responses can be used for data analysis
Probability sampling: a generalization of simple random sampling in which each unit in the population has a known (but not necessarily equal) and nonzero probability of selection. Probability sampling allows researchers to account for selection probabilities and make the same population inferences as with a simple random sample.
Nonprobability sampling: a family of sampling approaches that do not rely on known and nonzero sampling probabilities, but rather on implicit or subjective methods of selecting participants.

1.3 Random sampling error and sampling bias

Error in a statistical estimate, i.e. the difference between the true population parameter and the sample statistic, can stem from any stage of conducting a survey, particularly from measurement (e.g. through questionnaires), the sampling frame, the sampling method itself and from nonresponse. The goal of survey researchers is to minimize all sources of error in the final estimates. This perspective on the quality of a survey is sometimes called the *total survey error framework* (see Groves et al., 2004). It is important to note that in this framework, the concepts of random sampling error (or sampling variance) and sampling bias relate specifically to the sampling method used to select entities from the sampling frame into the gross sample. However, these concepts can also be considered foundational concepts of statistical inference.

The sampling error can be regarded as consisting of two parts: 1) the *random sampling error* stemming from *sampling variance*, which is the component that may over- or undershoot the population parameter for a single sample but is zero *on average* over many (hypothetical) repeated samples; the random sampling error thus shrinks with larger samples; and 2) the *sampling bias*, which is a *systematic* deviation that causes over- or underestimation even in large samples or on average over many (hypothetical) samples. We will describe the concepts of random sampling error and sampling bias in more detail below.

Selecting a (simple) random sample always leads to some degree of random difference between the sample statistic and the true population parameter. The extent of this difference is known as random sampling error. In classical (frequentist) statistical inference, a sample statistic is viewed as one possible outcome among many hypothetical repetitions of the same sampling process. To estimate how far the sample statistic might deviate from the true population parameter, statistical inference examines the distribution of these hypothetical outcomes, quantifying the sampling variance and thus the *expected* random sampling error.

The random sampling error is directly related to the sample size. If only few entities are randomly chosen as a sample, there is little chance that they will represent the population on average, i.e. any statistic estimated from this sample may be quite different from the population parameter it intends to capture. However, sampling more entities reduces the influence of random variations on the estimated summary statistics. It is possible to estimate the random sampling error from the values of a sample themselves, e.g. as the *standard error* of the sample mean, which indicates how much the sample statistic would vary across repeated samples from the same population. The standard error thus allows for quantifying the uncertainty of our estimates and enables tests of statistical *significance*.

The sampling bias is the more problematic part of the sampling error. It stems from a *systematically* (i.e. nonrandomly) unbalanced selection of entities from the population. For example, through their sampling process, researchers could be selecting a disproportionate number of young people even though they are interested in the entire population of a country. If the age of participants (or any other characteristic that may be misrepresented) is associated with the parameters researchers want to estimate, the estimates suffer from sampling bias.

Sampling bias cannot be prevented by choosing larger samples. Even if more and more entities are sampled in the same way, sampling bias cannot be expected to shrink. It is also impossible to directly recognize it from the collected values (e.g. the measurements of financial knowledge) themselves. In practice, there are different sources of bias that stem from different steps in the overall process of selecting entities from the population, as outlined in subsection 1.4.

Box 4

Key terms of error and bias

Random sampling error: the error in estimates that occurs due to random variation from sample statistics. This error decreases as the sample size increases.
Sampling bias: the systematic error of a statistic that is attributable to the sampling process. Unlike with the random sampling error, increasing sample size does not reduce the expected sampling bias.

1.4 Sources of bias

The overall bias from selecting survey participants can be differentiated into biases from different stages or from different processes in the overall sampling method. The following distinctions of biases are particularly relevant for probability sampling (see subsection 2.1), where samples are drawn in a systematic way. In nonprobability samples (see subsection 2.2), these distinctions may be artificial, for instance, if neither a clearly defined sampling frame nor a sampling procedure exist. However, this by no means implies that such biases are absent in nonprobability samples. Similar types of bias still occur, often even as a result of the chosen nonprobability sampling method. Moreover, the bias may be exceedingly more difficult to identify, categorize or correct due to the unstructured and opaque nature of nonprobability sampling.

First, it is relevant whether a reliable register of the population – a sampling frame – exists. Discrepancies between the target population and the sampling frame may lead to *coverage bias*. For example, the telephone book may be an inappropriate sampling frame for the target population of a city's residents and relying on it when sampling may lead to a large coverage bias. An official address register may be more suitable. However, it may lack residents who are not officially registered. Coverage can sometimes be improved by manually removing wrongly included respondents from the sample or sampling frame or by making an extra effort to include initially excluded respondents.

Second, the sampling method itself is important, i.e. the process of selecting entities from the sampling frame into the gross sample, e.g. the method of selecting individuals to be invited to participate. In probability sampling, the sampling process may be a simple random sample, which does not introduce *sampling bias*. Other, more complex approaches of probability sampling, however, may sample clusters of entities based on certain characteristics (cluster sampling, see subsection 2.1.3) or use adjusted sampling probabilities for certain groups of entities (stratified sampling, see subsection 2.1.2). In principle, these processes result in sampling bias through the sample design itself. Because in probability sampling, the sampling probabilities are known, this type of bias can be corrected through design weights (see subsection 4.1).

Nonresponse bias (also called participation bias) occurs when some of the contacted entities in the gross sample do not respond, which potentially leads to systematic differences between respondents and nonrespondents, i.e. between the gross and net samples. Nonresponse bias can be further differentiated: *Noncontact* bias may occur when people have been sampled but cannot be reached through the chosen contact method. *Self-selection* bias arises when contact is successful, but people choose not to participate. For example, it is possible that people interested in a survey's topic are more likely to participate (self-select) or that some people invited via e-mail do not regularly check their e-mail inbox (noncontact) – both of these scenarios can lead to nonresponse bias.

Box 5

Key terms of sources of bias

Coverage bias: bias stemming from discrepancies between the population and the sampling frame
Nonresponse bias (participation bias): a form of bias that arises when sampled entities in the gross sample do not respond to a survey. This may lead to systematic differences between respondents and nonrespondents, i.e. between the gross and net samples, and affect the outcome.
Noncontact bias: a subtype of nonresponse bias that occurs when entities cannot be reached with the chosen contact method
Self-selection bias: a subtype of nonresponse bias that occurs when entities opt in or out of a survey based on characteristics that are associated with the estimated outcome

1.5 Sampling for experimental impact assessments

To evaluate the impact of a financial education intervention, e.g. on the financial knowledge of school children, researchers may want to know what would happen if the population of children received the intervention in question. To this end, they need a control group, i.e. a sample from the population of school children that do not receive the intervention, and a treatment group, i.e. a sample from the (hypothetical) population of children that do receive the intervention. They may achieve this by drawing two random samples from the population of school children and providing financial education to only one of them.

This approach is largely equivalent to drawing a random sample from the population and subsequently randomly assigning individuals to treatment and control groups. Such random assignment is called *randomization*. It guarantees that the assignment process is independent of the outcome or any other sample characteristics. In principle, randomization allows clear causal interpretations of the treatment effects and guarantees internal validity.

However, randomization can also occur after an initial *nonrandom*, biased sampling of participants. For example, researchers could first draw a nonprobability sample of school students, e.g. using a process where schools self-select into the education program, and only then randomly allocate students to treatment and control groups. Indeed, such a process is often used in practice (see Lorenz et al., 2025). While it guarantees *internal* validity, it sacrifices external validity and generalizability. In other words: The difference between treatment and control groups can be clearly attributed to the education intervention itself. However, the treatment effect cannot be generalized to the whole population of school students unless one can assume homogenous treatment effects across the sampled and unsampled population.

Indeed, although many impact assessments of financial education interventions are based on randomized experimental designs, studies often appear to give little consideration to the sampling bias of treatment effects, even though it appears likely that these effects would differ between the sampled and nonsampled parts of the population, e.g. in cases when schools self-select into studies on financial education interventions (see Lorenz et al., 2025).

2 Sampling methods

Although the simple random sample can be regarded as the foundation of quantitative research and statistics, the different goals and practical limitations of studies necessitate a range of other (nonrandom) sampling approaches. Sampling methods can be broadly categorized into probability and nonprobability sampling techniques.

Probability sampling techniques include all generalizations of simple random sampling that still employ random selection processes, even if some groups may have larger or smaller probabilities of being sampled than they would have in a simple random sample. In probability sampling, every entity of the population frame has a *known* and *nonzero* probability of being selected. Using weighting methods, the sampling probabilities can be counterbalanced for statistical analyses (see subsection 4.1).

In a nonprobability sample, nonrandom and often subjective methods are used to collect the sample, which means that the probability of selecting entities of the population is *unknown* or *zero*. While nonprobability sampling approaches can be valuable in the appropriate research contexts, they differ considerably from probability sampling and are typically severely limited in their ability to generalize findings to the entire population.

Understanding common sampling approaches may not only help in choosing a sampling approach for collecting data, but also in interpreting their results and the relevance they hold for policymakers and practitioners in the field of financial education. For instance, different sampling approaches may at least in part explain potential discrepancies in the results of different surveys on the same issues. The following subsections aim to provide a general overview of the key types of probability sampling, including a general outline of more complex approaches such as multistage stratified sampling.

2.1 Probability sampling

Probability sampling describes selection processes based on objective probabilities rather than subjective decisions. The simplest form of probability sampling is a simple random sample (as outlined in subsection 1.4). In a simple random sample, every entity in the population has the same independent chance of being selected. Other probability samples are generalized forms of the simple random sample, where each element in the target population has a *known* and *nonzero* chance of being selected into the sample.

Under ideal circumstances (i.e. a perfect sampling frame and no nonresponse), probability samples combined with proper design weights (see subsection 4.1) enable statistical estimates without sampling bias. Although probability sampling allows bias-free estimates in principle, in practice all sampling approaches tend to introduce bias. While weighting may help to reduce bias, it rarely eliminates it entirely (see section 4). Nevertheless, probability samples must be considered superior to nonprobability samples (Cornesse et al., 2020).

Probability sampling of large populations tends to be costly. It requires a suitable sampling frame, an appropriate method of contacting and incentivizing participants (see subsection 3.1) and considerable organizational effort and methodological expertise. Different methodological approaches can reduce the organizational effort and cost of probability sampling, as with cluster sampling (see section 2.1.4) for instance, but increase the complexity of the sampling and weighting process.

2.1.1 Simple random sampling

As outlined in section 1, simple random sampling is the most fundamental form of probability sampling and serves as the basis for statistical inference. In simple random sampling, every unit of the target population has an equal probability of being selected. In practice, a computer randomly selects the desired number of units from the sampling frame.

A nuance in (simple) random sampling is the distinction between sampling with replacement and sampling without replacement. Sampling with replacement would mean that any element of the population is replaced into the sampling frame after being selected and can be sampled more than once. As it is typically not desired that the same individuals participate in one and the same survey several times, sampling without replacement is used in practice (Groves et al. 2004).

2.1.2 Stratified sampling

Sometimes researchers want to guarantee representativeness with regard to some known characteristics of the population, e.g. residency in a geographical region. If the relevant characteristic is known and available to be included in the sampling frame, the sampling frame can be divided into subgroups based on this characteristic, and random samples are drawn from each subgroup. This ensures that the subgroups of interest are all well represented in the sample. The subgroups are called *strata* and the sampling approach is called *stratified sampling*.

Stratification protects against obtaining a sample that does not match known distributions of characteristics. Moreover, stratification can guarantee precise estimates for the specific strata. For example, in a national survey, researchers may want to ensure that both urban and rural populations from different regions of the country in question are represented adequately. A simple random sample might make it difficult to obtain precise estimates for some rural areas. By stratifying the sample on the basis of urban and rural regions, researchers can guarantee that both groups are properly covered and that accurate comparisons between them will be possible.

Sample sizes within the strata can be adjusted to match the intentions of the researchers. For instance, with proportional sampling, the sample size from each stratum proportionally reflects the size of the respective subpopulations. Using disproportionate sampling, the sample size in each stratum could be adjusted to achieve the required precisions of estimates. To achieve this purpose, there are techniques for the optimal allocation of sample sizes to strata. Based on prior estimates of variability within strata, these techniques can maximize statistical efficiency (e.g. the Neyman Allocation, Neyman, 1934; see also Groves et al., 2004).

2.1.3 Cluster sampling

In this section, we have assumed, so far, that researchers can obtain a suitable sampling frame that contains all relevant population members and that there are no practical limitations to the sampling process and the field work (i.e. recruitment and survey modes). In reality, however, both the sampling process and the fieldwork may be limited due to data availability and financial or time resources. To facilitate the sampling process and field work, probability sampling can be based on sampling groups (so-called *clusters*) of entities rather than on the entities themselves.

For example, researchers may want to obtain a probability sample from the population of all school students of a certain age to test their financial literacy (e.g. as in the OECD's PISA assessments, see OECD, 2024). However, it may be unfeasible or undesirable to 1) obtain a sampling frame of all eligible students, 2) directly contact students or 3) organize tests for students from too many schools. Researchers may thus decide, instead, to sample groups of students that are naturally organized in *clusters*.

For instance, researchers could draw a simple random sample of *schools* as clusters of students, instead of directly drawing a simple random sample of students from all schools. In this way, a probability sample is still possible, because 1) all students have a nonzero probability of being sampled via their school cluster and 2) the sampling probability of each sampled student can be calculated from the number of students per sampled school. Sampling a subset of schools can also facilitate the recruitment and testing procedures, because researchers may be able to rely on organizational support by the sampled schools' administrations and teachers.

Clusters may also be geographical regions. For instance, instead of sampling from the population of all individuals in a country to assess financial literacy levels, researchers may draw a sample of geographic administrative districts. Geographic clusters can considerably reduce organizational and travel costs of in-person interviews, because the sampled clusters of participants are located closely together and interviewers can visit them more easily and quickly.

Importantly, clusters may be populated by too many entities to interview or test all within a sampled cluster. This means a second sampling stage within clusters may be necessary, as outlined in subsection 2.1.4.

2.1.4 Multi-stage sampling

Multi-stage sampling involves selecting samples through a series of stages, moving from larger groups to progressively smaller ones. In practice, this often involves clustered sampling, where the population is first divided into clusters and sampling is then carried out at different levels of aggregation.

In the example of sampled regional clusters given in subsection 2.1.3, we noted that clusters may contain too many students to allow for interviewing all of them within each sampled cluster. Researchers may thus turn to another sampling stage at the level of individuals within clusters, which results in a two-staged clustered sampling design. For example, in stage one, ten schools (clusters) may be randomly selected. In stage two, within each selected school, twenty students are randomly sampled to participate in the survey. Indeed, the OECD's PISA assessments use a second sampling stage at the student level (OECD, 2024).

In the case of regional clusters, researchers may sample a certain number of individuals from each sampled cluster, e.g. a fixed number of participants or a number that is proportional to the subpopulation size of the regional cluster. To obtain a true probability sample, this second stage requires a complete sampling frame of individuals within clusters to enable sampling.

Multi-stage sampling can also be combined with stratified sampling. For instance, stratification may occur at the level of large geographical regions to ensure that these regions are well represented in the sample, while clustering happens at the level of smaller geographical regions for practical reasons and to reduce costs and final sampling at the level of individuals to limit the number of total participants. This design would be referred to as a stratified, multistage sampling design with clustering at the second stage.

In practice, the final sampling stage is sometimes quasi-probabilistic. For example, a survey could rely on *random route patterns*, i.e. interviewers following a pre-defined randomized path, starting from a randomly selected point within a given area. While such sampling methods may be necessary, for instance, if suitable sampling frames are unavailable, they are not considered probability sampling and may lead to bias (see e.g. Bauer, 2014). In other approaches, households are sampled first, and then individuals within households are selected by using probability or quasi-probability approaches, e.g. the Kish method or the birthday method, which may influence bias and interview refusals in different ways (see e.g. Jabkowski, 2017).

In survey research, sampling levels are typically numbered from the most aggregate level downward: starting with primary sampling units (PSUs) that are followed by secondary sampling units (SSUs), tertiary sampling units (TSUs) etc., down to the ultimate sampling units (USUs), which typically represent the entities of interest as defined by the target population.

As long as the sampling probabilities from the different sampling stages are known for the USUs (e.g. the individuals), design weights can be calculated that offset the different probabilities of individuals being selected into the sample. We give a short overview of weighting in section 4.

2.2 Nonprobability sampling

Nonprobability sampling is key to *qualitative* empirical studies that do not aim at the generalization of findings in a statistical sense but rather at the in-depth exploration of a phenomenon. However, non-probability sampling is also frequently used with quantitative surveys for the reasons outlined below.

First, nonprobability sampling typically costs less than probability sampling. Probability samples are more expensive as they require more rigorous planning, complete contact data and the contacting of entities (often multiple times). The economic downside of probability sampling has become more pronounced recently as these methods often face low response rates (Elliott and Valliant, 2017). In contrast, nonprobability sampling typically relies, for instance, on volunteers, people living in geographical proximity, personal acquaintances, colleagues or members of an institution such as university students, which makes it easier to gather a substantial number of responses. While it may be economically viable, it is important to recognize that nonprobability sampling differs fundamentally from probability sampling in terms of the ability to make inferences about the target population.

Second, experimental studies such as randomized controlled trials often rely on nonprobability sampling. Although randomization ensures internal validity, generalizing results to a broader population rests on the assumption that the treatment effects are *homogeneous* across all units within the relevant population – an assumption that may not always hold. Randomization in the assignment of participants to treatment and control groups must therefore be distinguished from random or probability sampling of study participants (see subsection 1.5).

Third, certain research interests focus on populations for which comprehensive registers or lists are unavailable, i.e. where the population itself is not perfectly known. In financial education research, such populations may include groups like migrant communities, the unemployed or individuals at risk of poverty. For such populations, probability sampling may not be feasible and statistical representativeness may not be the primary research concern. Instead, the goal is to gain insights into specific phenomena or key issues affecting certain groups (Statistics Canada, 2021). In the following, we briefly present examples of nonprobability sampling techniques (for more details, see e.g. Statistics Canada, 2021; Vehovar et al., 2016).

2.2.1 Convenience sampling

Convenience sampling is a nonprobability sampling technique where participants are selected on the basis of their availability and proximity to the researcher, without any further selection criteria being applied. Typically, this method involves limited planning and preparation, which makes it a convenient approach to data collection. Participants may, for example, be recruited through personal contacts, through institutions or by approaching individuals in public (online) spaces. Convenience sampling is useful for gathering sizable data within a short period; the results it yields, however, are prone to severe sampling and nonresponse bias.

2.2.2 Volunteer sampling

Volunteer sampling relies on nonpersonalized invitations, thus allowing entities from the target population to actively opt into the sample. This approach emphasizes the proactive effort required by participants to self-select into the sample. Examples of volunteer sampling include invitations to a large audience (e.g. on TV, the radio or social media) or the use of printed invitations (e.g. posters or leaflets). An issue that is particularly pronounced with volunteer sampling is the problem of self-selection bias (see subsection 1.4). Those who actively volunteer to participate in the survey may differ significantly from those who do not (see Heckman, 1990).

2.2.3 Judgment sampling

Judgment sampling uses specific criteria or expert insights to actively select participants considered relevant to the research topic, aiming for a context-specific form of "representativeness" rather than statistical generalizability. Judgment sampling is useful when the composition of the target population is unknown or when it is defined by an abstract concept, such as "people who are financially vulnerable." In such cases, researchers establish suitable selection criteria and actively recruit participants accordingly.

Judgment sampling can be susceptible to subjective biases if the researcher has an inaccurate understanding of the target population. Therefore, the theoretical basis of the sample selection should be made transparent.

2.2.4 Quota sampling

Quota sampling is often used in convenience or volunteer samples or in web panels. It aims to align sample characteristics with the distribution of certain sociodemographic characteristics found in the target population, often including e.g. age, gender and education. Researchers thus select participants in a way that the resulting sample fulfills the quotas derived from census data. In practice, this may be achieved through the recruitment process, e.g. by distributing invitations through channels relevant for certain groups, or by simply rejecting people from participating in a survey if a quota is already fulfilled.

Although quota sampling utilizes information about the target population and may superficially resemble stratified sampling (see subsection chapter 2.1.3), it remains a nonprobability sampling method. As such, despite its ability to match certain characteristics of the target population, it likely suffers from (unquantifiable) sampling bias. In market research and opinion polling, quota samples may sometimes be labeled "representative," as they match demographic quotas of the population (e.g. all age groups may be well "represented"). However, they cannot be called statistically representative, as they do not meet the criteria for statistical inference.

2.2.5 Access panels

Online access panels (which are sometimes called "online panels" or "web panels") are a pool of registered participants who expressed a willingness to participate in future surveys. In response to declining participation rates in traditional surveys, nonprobability access panels have become widely used as cost-effective alternatives (Elliott and Valliant, 2017). While they are widely utilized not only in market research and opinion polling but also increasingly in scientific studies, access panels typically rely on volunteer sampling and financial incentives to achieve high participation rates. They should thus not be viewed as a replacement for probability sampling in studies where statistical representativeness is critical.

With access panels, polling firms maintain a pool of registered respondents who self-select into various surveys. Within these panels, quota sampling is often employed to align the sociodemographic characteristics of the sample with those of the target population (see subsection 2.2.4). Access panels must not be confused with probability-based panel designs such as rotating panels, which aim to collect longitudinal data from probability samples over multiple points in time.

3 Fieldwork

After selecting the sample, the next steps are known as fieldwork. Fieldwork includes recruiting participants and collecting data through the chosen survey modes. Fieldwork may be closely linked to bias, particularly with regard to nonresponse bias, as people's willingness to participate in a survey when contacted can be associated with both the survey mode and the recruitment strategy. Below, we provide an overview of recruitment strategies and survey modes.

3.1 Recruitment strategies

The goal of recruitment is to reach as many sampled entities as possible to maximize participation and minimize survey nonresponses. However, recruitment strategies depend largely on the available contact information for potential survey participants and vice versa. For instance, the available sampling frame may only include postal or e-mail addresses, which limits the potential contact channels.

When researchers can choose among different contact channels, they should carefully consider how certain contact channels are perceived by potential participants and how they affect response rates. For example, e-mails may often go unread or raise concerns about their trustworthiness. In contrast, personal visits before the interview or invitation letters tend to be more effective, yet perhaps costlier, alternatives (Daikeler et al., 2020).

Prenotification is a method to provide advance notice to respondents before recruitment. It has been studied as a potential strategy to increase participation. However, research suggests that prenotification is only effective when traditional communication channels such as letters or phone calls are used. In web-based surveys, prenotification efforts may have limited impact, as e-mails are more likely to be ignored (Daikeler et al., 2020).

Providing *incentives* is another effective strategy used to reduce nonresponses, with monetary incentives being the most effective one (Abdelazeem et al., 2023). Monetary incentives can be unconditional (paid before survey participation) or conditional (paid after survey completion). Unconditional incentives have been shown to be more effective, as participants may feel a sense of obligation to complete the survey in exchange for the monetary reward (see Abdelazeem et al., 2023, for a meta-study on the effectiveness of monetary incentives to increase response rates).

For some nonprobability sampling methods, sampling and recruitment occur simultaneously, as sampling is directly tied to participant recruitment. For example, in convenience sampling via a social media announcement, individuals are both selected and invited to participate at the same time. In such cases, the mode of data collection (discussed in detail below) may be directly linked to recruitment: When participants are recruited online, web-based surveys are likely to be much more convenient than arranging phone or face-to-face interviews.

Although probability sampling typically follows distinct procedures that can be considered independent, practical synergies between sampling, recruitment and survey mode can be used. For instance, if a personal interview at the respondent's home is planned, sending an advance prenotification letter to a sampled address to announce the interviewer's visit may increase participation. The actual contact attempt by an interviewer should then lead to the interview, if possible. Similarly, for telephone interviews, immediate recruitment via telephone may be the most practical approach.

3.2 Survey modes

Once participants agree to take part in a survey, their responses are collected using different survey modes (also referred to as interview modes). These modes can be categorized based on two key factors: The first factor is the data collection *channel*, which determines whether the survey is conducted face-to-face, by telephone, online or through a paper-based format. The second factor is the type of *administration*, which differentiates between interviewer-administered and self-administered surveys.

In interviewer-administered surveys, an interviewer poses the questions and records the responses. In contrast, self-administered surveys require respondents to complete the questionnaire independently. Face-to-face and telephone surveys are typically interviewer-administered, while web-based and paper-based surveys are self-administered.

A growing trend in survey methodology is the use of mixed-mode surveys, where multiple modes are combined. This approach may help mitigate biases associated with self-selection or the undercoverage of certain population groups, as it allows respondents to participate through different means while it potentially also reduces costs (Revilla, 2010). For example, older populations with limited internet access may be underrepresented in web-based surveys, while younger individuals may be less inclined to engage in face-to-face or telephone interviews. By integrating multiple modes or allowing respondents to select their preferred method, researchers may be able to reduce certain forms of bias. Additionally, mixed-mode surveys improve accessibility when some segments of the population require assistance in completing the survey.

Most survey modes utilize digital tools to facilitate data collection. Three common digital-assisted methods are *computer-assisted personal interviewing* (CAPI), *computer-assisted telephone interviewing* (CATI) and *computer-assisted web interviewing* (CAWI). In CAPI, the interviewer conducts the survey in person while entering responses into a digital device. In CATI, the interviewer administers the survey over the telephone while simultaneously recording responses digitally. Different types of online surveys are generally labeled CAWI.

The choice of survey mode depends on two primary considerations: *mode effects* and *resource constraints*. Mode effects refer to differences in *how* respondents respond and *which individuals participate* in the survey given the survey mode. These differences can arise from how questions are presented, how comfortable respondents feel providing answers or from the extent of interviewer involvement (Groves et al., 2004). Mode effects, therefore, play a crucial role in shaping data quality.

Self-administered surveys tend to reduce *social desirability bias*, which may make them particularly suitable for sensitive topics. However, they also tend to have higher *item non-response rates* or implausible responses as respondents may skip certain questions or may rush through the survey. An interviewer, on the other hand, can help clarify ambiguous questions, encourage participation and increase response rates. However, interviewer presence can also influence participants' responses, a phenomenon known as *interviewer effect*. This effect is especially pronounced when there is a significant power or authority gap between the interviewer and the respondent, such as an adult interviewing a child (Vogl, 2021).

Another key aspect of mode effects is whether survey questions are presented *verbally* or *visually*. Groves et al. (2004) explore the effects of some survey question presentation effects and conclude that verbal surveys, such as those conducted by an interviewer, often lead to *recency effects*, where respondents are more likely to choose the last response option they hear. In contrast, Groves et al. (2004) note that visually presented survey questions, such as those in web-based or paper-based formats, are more susceptible to *primacy effects*, where respondents tend to select the first option they see. The way questions are displayed also influences response patterns and can lead to specific biases. For example, *straight-lining bias* occurs when respondents repeatedly select the same answer category; *central tendency bias* refers to the tendency to avoid extreme responses and consistently select middle answer categories (see Groves et al., 2004).

Different survey modes are associated with varying response rates. Meta-studies indicate that face-to-face interviews consistently yield the highest response rates, while web-based surveys experience the lowest (e.g. Daikeler et al., 2020). These differences are largely attributed to variations in recruitment strategies. When respondents are approached by an interviewer they perceive as credible they may be more likely to participate in the survey. In contrast, web-based surveys often rely on e-mail invitations, which may be disregarded (see subsection 3.1). The COVID-19 pandemic also appears to have had a lasting impact on survey participation. Since the pandemic, response rates – in particular in face-to-face interviews – appear to have declined significantly – a fact that has contributed to an increase in mixed-mode surveys (Charman, 2024; European Social Survey, 2025a).

For instance, the recent Austrian wave of the OECD's International Survey of Adult Financial Literacy (see Voith and Zieser, 2024a, 2024b) combined around 30% of CAWI with 70% of CAPI. Respondents were randomly assigned to either mode. Prenotification letters were sent to participants in both modes. While CAPI participants did not receive incentives, CAWI respondents received a EUR 5 unconditional incentive included in the invitation letter, followed by an additional EUR 10 upon survey completion.

Despite these incentives for CAWI respondents, response rates remained notably lower, with CAPI achieving an average response rate of approximately 40%, compared to only approximately 20% for CAWI.

While the literature provides only a limited number of precise quantifications of economic costs across survey modes, web-based interviews are generally associated with lower financial costs, whereas face-to-face interviews are the most financially demanding due to their time- and labor-intensive nature. The costs of in-person interviews can be particularly high when experienced interviewers are required to travel long distances, make multiple contact attempts or conduct follow-up visits.

Choosing the most appropriate survey mode can be complex as it involves balancing mode effects, cost structures and survey objectives. Each mode presents unique advantages and disadvantages, making no single method universally superior. Researchers must carefully weigh trade-offs based on their study's priorities.

To minimize bias, face-to-face interviews – which can be supplemented by web-based interviews – continue to be recommended (see Daikeler et al., 2020). Due to their high cost, they are typically used for large-scale scientific surveys by public authorities that demand high standards. For highly sensitive topics, self-administered surveys can yield better data quality as they minimize social desirability bias. Particularly when surveying children, interviewer-administered modes may introduce authority bias, making self-administered surveys more suitable. However, a trade-off between sensitivity and complexity often needs to be managed: for example, actual financial outcomes or behaviors are likely to be highly sensitive, while their composition can be complex and difficult for non-experts to follow. In such cases, the presence of trained interviewers can help improve data quality by assisting respondents in navigating complex questions.

For studies involving repeated cross-sectional or panel data, it is crucial to maintain mode consistency over time. Frequent changes in survey mode can introduce measurement inconsistencies, making longitudinal comparisons unreliable. Thus, stability in the chosen survey mode is essential to ensure data comparability across survey waves.

4 Weighting

Weighting aims to reduce the sampling bias of estimated statistics that stems from unequal selection probabilities in complex probability sampling approaches by applying *design weights* and nonresponse bias by applying *nonresponse weights.* Moreover, under certain conditions, *post-stratification* or *calibration weights* can further reduce random sampling error and nonresponse bias by adjusting sample weights so that the distribution of some of the sample's characteristics match known distributions of these characteristics in the population. In practice, different types of weighting approaches can be combined.

It is important to note that weights typically increase the *variance* and *standard error* of the estimated statistics. Variance can increase because weighting gives more influence to entities from groups that are underrepresented in the sample compared to the population. Moreover, responses from sampling units within clusters are often correlated. For example, students sampled within a subset of schools may share similar characteristics. This *intracluster correlation* reduces the effective sample size because the information gained from each additional entities within a cluster adds less information than the information gained from entities that were sampled independently.

While weights can mitigate different sources of bias, complex sampling designs can reduce the precision of estimates compared to cases in which a simple random sample of the same size is employed. For accurate variance estimation and statistical inference, the sampling design and data structure must be incorporated into data analysis to obtain valid variance estimates, standard errors, confidence intervals and significance tests. Methods of variance estimation are beyond the scope of this paper but can be found e.g. in Groves et al. (2004), Lohr (2021) or ECB (2023).

4.1 Design weights

Design weights directly offset differences in sampling probabilities among entities of the sampling frame that result from complex probability sampling methods such as stratified or clustered sampling. For instance, if a fixed number of individuals is sampled from regional strata (see subsection 2.1.3), individuals from a more populated region have a smaller probability of being sampled than individuals from a less populated region. Without accounting for this circumstance, the individuals from the more populated region would be underrepresented in the sample. Similarly, if in a two-stage design individuals are sampled from households at the second stage, individuals from one-person households have a larger probability of being sampled than individuals from multi-person households.

A way to counteract these unequal probabilities of being selected into the gross sample is to calculate the different sampling probabilities within strata and clusters. Using the reciprocals of the sampling probabilities as weights for the sampled units of analysis (e.g. individuals) corrects for unequal selection probabilities and offsets the corresponding bias in the statistics.

Let us suppose, for example, that we use a stratified two-stage sampling approach where we sample 100 households for each of ten regions as strata. This is a disproportionate stratified sampling approach because the strata are populated by different numbers of households. From each household, we sample one individual, regardless of the size of the household. The sampling probability for one sampled individual from a two-person household in a region with 100,000 households may thus be calculated as follows: 1) The individual was sampled from a two-person household, i.e. as one from two possible individuals, with a probability of 1/2. 2) The household was sampled as one of 100 from a region with 100,000 households with the probability 100/100,000. The overall sampling probability of the individual is the product of these two probabilities, namely 1/2,000, and its weight is 2,000. Weights calculated in this fashion can be applied to the calculation of statistics (e.g. point estimates such as means or proportions).

4.2 Nonresponse weights

Nonresponse weights aim to reduce nonresponse bias, i.e. bias that stems from an underrepresentation of participants that were sampled in principle (i.e. who were included in the gross sample) but could not be contacted or refused to participate (i.e. who were not included in the net sample). If nonresponse is associated with characteristics that are known for both sampled participants and sampled nonparticipants (i.e. for the entire gross sample), the probability of participation can be predicted for each entity on the basis of these characteristics.

The characteristics used as predictors may not only stem from information about the individuals themselves, but also from *paradata* such as interviewer characteristics, neighborhood observations or the documentation of contact attempts noted by the interviewer (Olson, 2013). The reciprocals of these participation probabilities are then the basis for the nonresponse weights that are applied each of the participants.

Suppose, for instance, that researchers have information about the gender, age and zip code of all sampled individuals. After sampling and data collection, and now knowing which sampled individuals have actually participated, they can use statistical models to predict the probability that an individual participates based on their gender, age and zip code (and any combinations thereof). The inverse of this probability can serve as the basis for the nonresponse weight for those that actually did participate.

For instance, young men living in large cities may, on average, be comparatively hard to contact or likely to refuse participation in an interview. Those who *do* participate are assigned low response probabilities based on their characteristics, and thus high nonresponse weights. Data from young men living in cities is thus upweighted so that they are better represented in the population estimates and nonresponse bias is reduced. This logic is of course applied to all participants and combinations of characteristics in this hypothetical example.

4.3 Post-stratification and calibration weights

Post-stratification is usually the last step in the weighting procedure. It aims to weight a sample so that its characteristics match the known distributions of these characteristics in the target population. It is thus similar to stratified sampling (see subsection 2.1.2) in its purpose, only that post-stratification is implemented after sampling and data collection.

For example, suppose researchers know (from reliable census data or other sources) that 70% of the target population live in region A and 30% in region B. However, in their sample, 50% of participants live in region A and 50% in region B. Decreasing the weights for all participants living in region A and increasing them for participants living in region B so that the sample's regional distribution matches the known population distribution ensures that both regions are properly represented in estimates and may help reduce bias.

If the characteristics used are correlated with the outcomes of interest (e.g. the region in question with the respective digital financial literacy scores), post-stratification can help reduce both nonresponse bias and random sampling error. For example, individuals from urban regions may be slightly more digitally financially literate than individuals from rural regions. If urban respondents are underrepresented, e.g. due to self-selection, the overall estimate of the average digital financial literacy of the population may be slightly underestimated. This bias can be reduced through post-stratification by region.

There are extensions of post-stratification (e.g. regression-based post-stratification methods and *calibration* methods) where the distribution of (sometimes complex combinations of) multiple variables is used to minimize the overall deviation between sample characteristics and known population distributions. If the necessary variables are collected in the survey, such approaches can be used e.g. with distributions of age, gender, geographic region and other information available from national censuses or other reliable sources.

Post-stratification or any of its extensions thus hinge not only on the accuracy and availability of the benchmark distributions but also on the strength of the association between the characteristics used and the outcomes of interest. Even when many reliable benchmark distributions are used, post-stratification cannot correct for all sources of bias if relevant predictors of the outcome are omitted.

5 Conclusion

This paper provides an accessible overview of key components of survey methodology, with a particular focus on their relevance for achieving findings in research on financial literacy and financial education that can be generalized to the intended populations. Outlining fundamental methodological principles in an applied research context, it seeks to support both researchers and program designers in making informed decisions throughout the research process. It also aims to raise awareness of how methodological design influences the generalizability and the interpretability of findings.

Section 1 lays the theoretical foundation of the paper by introducing central concepts such as representativeness, target population, sampling frame, sampling error and biases. These concepts provide a rationale for why sampling is critical when it comes to statistical inference and to generalizing results from impact assessments and large-scale national surveys to broader populations. The subsequent sections relate these conceptual foundations to practical methods of sampling, fieldwork and weighting.

Section 2 emphasizes the importance of choosing appropriate sampling strategies to ensure that study results meaningfully reflect broader populations. It distinguishes between probability and nonprobability approaches and outlines widely used sampling strategies such as stratified, clustered and multi-stage sampling.

Section 3 focuses on fieldwork, showing different recruitment strategies and survey modes. It discusses how the choice of contact methods, incentives and data collection modes (e.g. self-administered vs. interviewer-administered; single-mode vs. mixed-mode) can significantly influence response rates and introduce or mitigate bias.

Section 4 introduces weighting procedures as tools to correct for various biases from the sampling method and from nonresponse. It explains the rationale behind design weights, replicate weights and post-stratification.

This paper is limited in scope and provides only a brief conceptual foundation to a vast and increasingly complex field of survey methodology, focusing on the most fundamental available methods and simplifying certain complexities. For instance, it does not address advanced sampling designs for longitudinal surveys, such as rotating panel designs (see Lynn, 2009), and it does not explore item-level nonresponse or imputation methods (see Little and Rubin, 2020). Also, the development of survey instruments themselves – such as selecting question types, testing for item validity and reliability, the importance of interviewer training and pretesting – goes beyond the scope of this paper.

Nonetheless, by condensing and simplifying complex issues, this paper helps newcomers and non-specialists deal with surveys in a financial literacy context. We also hope to contribute to a deeper understanding of the genesis of survey data by highlighting the importance of survey methodology as a sometimes neglected aspect that is essential to ensuring the quality of research in financial literacy and education.

References

Abdelazeem, B., A. Hamdallah, M. A. Rizk, K. S. Abbas, N. A. El-Shahat, N. Manasrah, M. R. Mostafa and M. Eltobgy, M. 2023. Does usage of monetary incentive impact the involvement in surveys? A systematic review and meta-analysis of 46 randomized controlled trials. In: PloS one 18(1). e0279128.

Albacete, N., P. Lindner and K. Wagner. 2023. Eurosystem Household Finance and Consumption Survey 2021. Methodological notes for Austria. OeNB Reports 2023/1. OeNB. https://www.oenb.at/dam/jcr:eecdfa0e-ac26-4a62-a378-90142549e410/HFCS-Austria-2021_Addendum_Methodological-Notes.pdf

Bauer, J. J. 2014. Selection errors of random route samples. In: Sociological Methods & Research 43(3). 519–544.

Charman, C., S. Mesplie-Cowan and D. Collins. 2024. The post-pandemic role of face-to-face fieldworker. National Centre for Social Research. https://natcen.ac.uk/sites/default/files/2024-02/Survey%20future%20report%20v4.1.pdf

Cornesse, C., A. G. Blom, D. Dutwin, J. A. Krosnick, E. D. De Leeuw, S. Legleye, J. Pasek, D. Pennay, B. Phillips, J. W. Sakshaug, B. Struminskaya and A. Wenz. 2020. A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. In: Journal of Survey Statistics and Methodology 8(1). 4–36.

Daikeler, J., M. Bošnjak and K. L. Manfreda. 2020. Web versus other survey modes: An updated and extended meta-analysis comparing response rates. In: Journal of Survey Statistics and Methodology 8(3). 513–539.

European Central Bank. 2023. Household Finance and Consumption Survey: Methodological report for the 2021 wave. In: ECB Statistics Paper Series No 45.

Elliott, M. R. and R. Valliant. 2017. Inference for nonprobability samples. In: Statistical Science 32(2). 249–264.

European Social Survey. 2025a. Data collection: Switch to self-completion. https://www. europeansocialsurvey.org/methodology/methodology/data-collection-sc

European Social Survey. 2025b. Methodology overview. https://www.europeansocialsurvey.org/methodology/methodology-overview

Groves, R. M., F. J. Fowler Jr., M. P. Couper, J. M. Lepkowski, E. Singer and R. Tourangeau. 2004. Survey methodology. John Wiley & Sons.

Heckman, J. J. 1990. Selection bias and self-selection. In: Eatwell J., M. Milgate and P. Newman. (eds.). Econometrics. Springer. 201–224.

Jabkowski, P. 2017. A meta-analysis of within-household selection impact on survey outcome rates, demographic representation and sample quality in the European Social Survey. In: Ask: Research and Methods 26. 31–60.

Little, R. J. A. and D. B. Rubin. 2020. Statistical Analysis with Missing Data (3rd edition). Wiley.

Lohr, S. L. 2021. Sampling: Design and analysis. Duxbury Press.

Lorenz, T., S. Mauser and M. Zieser. 2025. Is financial knowledge enough? Reviewing impact assessments of financial education interventions. OeNB Financial Literacy Evaluation Series.

Lynn, P. (ed.). 2009. Methodology of Longitudinal Surveys. Wiley.

Neyman, J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. In: Journal of the Royal Statistical Society 97(4). 558–625.

OECD. 2023a. OECD/INFE 2023 international survey of adult financial literacy. OECD. https://doi.org/10.1787/56003a32-en

OECD. 2024. PISA 2022 Technical Report. OECD. https://doi.org/10.1787/01820d6d-en

Olson, K. 2013. Paradata for nonresponse adjustment. In: The Annals of the American Academy of Political and Social Science 645(1). 142–170.

Revilla, M. 2010. Quality in unimode and mixed-mode designs: A multitrait-multimethod approach. In: Survey Research Methods 4(3). 151–164.

Statistics Canada. 2021. Statistics: Power from data! https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm

Vehovar, V., V. Toepoel and S. Steinmetz. 2016. Non-probability sampling. In: Wolf, C., D. Joye, T. Smith and Y. Fu. (eds.). The sage handbook of survey methodology. SAGE.

Vogl, S. 2021. Mit Kindern Interviews führen: Ein praxisorientierter Überblick. In: Hedderich, I., J. Reppin and C. Butschi. (eds.): Perspektiven auf Vielfalt in der frühen Kindheit. Mit Kindern Diversität erforschen (2nd edition). Verlag Julius Klinkhardt. 142–157.

Voith, V. and M. Zieser. 2024a. International Survey of Adult Financial Literacy 2023: first results for Austria. OeNB Reports 2024/13. OeNB. https://www.oenb.at/dam/jcr:ce083d11-f899-4dc5-bdb5-43-a32a537b84/report-2024-13-financial-literacy.pdf

Voith, V. and M. Zieser. 2024b. OeNB Barometer: methodological notes 2023. OeNB Report 2024/14. OeNB. https://www.oenb.at/dam/jcr:96a02d40-7d88-4962-8e98-e3dd86261c0b/report-2024-14-oenb-barometer-methodological-notes-2023.pdf